Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/110119

Bayesian Parameter Estimation and Model Selection for Gallbladder Cancer Data of two Countries

Richa Srivastava[1,*], *Rakesh Ranjan*[2] *and Himanshu Misra* [1]

[1]Jaipuria Institute of Management, Lucknow, India
[2]DST-Center of Interdisciplinary Mathematical Sciences (CIMS), Banaras Hindu University, India

Received: 14 Mar. 2021, Revised: 29 Jun. 2021, Accepted: 2 Jul. 2021
Published online: 1 Jan. 2022

Abstract: The paper proposes statistical model and complete Bayesian inference for cancer survival data of two countries. Complete posterior analysis is done by generating random samples from posterior surface. Gibbs sampler, a Markov chain Monte Carlo (MCMC) method has been used, for generating samples from posterior distribution. The paper also provides algorithm for Gibbs sampler generation scheme for proposed model parameters as well its density estimation. Model compatibility and inter model comparisons, using the measures of Bayesian information criterion (BIC) and deviance information criterion (DIC) has been used.

Keywords: Bayesian Inference, Gibbs Sampler, BIC, DIC, Gallbladder carcinoma

1 Introduction

In the recent past, there has been drastic increase in the application of Bayesian methods. Now a days Bayesian methods have also been used in medical sciences due to its wide applicability and advanced computation methods. Although the conceptual appeal of Bayesian methods has been well understood by engineers, management, social sciences, demographer yet its applicability in medical field is less documented. Most often, only observed data is being used in medical research. Medical science is very sensitive field where it is very tedious to obtain precise information from patients suffering from any malignant or a very rare disease. As a result of this, it becomes very difficult to suggest new and appropriate methods of treatment of particulate disease as well as analyzing the after effect of the same.
According to Bayesian concept apart from observed data, information can also be captured from the experts of the field i.e. doctors, surgeons and other practitioners. Bayes concept motivates the use of information present with experts. In usual practice as well, we use the experience of an expert of the particular domain. These experienced persons may be a scientist, a professor, an engineer, a well known surgeon/doctors (see, Srivastava et al. (2018)). However the main issue is to combine information coming form the two sources, one from the expert of the area, and second from the observed data. In simple words, we need statistical tools which enable us to incorporate the flavor of experience of an expert into the statistical inference of the scenario because we can never ignore the opinion/judgment or expertise of well experienced person (see, O'Hagan (1998), Berger (2013)).
Application of statistical techniques to medical data problems also depends on the type of data. The medical data may be made available in a variety of ways. One such availability may be in the form of counts. An example can be case-control (Bassetti et al. (2004)). In this study, there may also be information on covariates or also referred sometimes as etiological factors. To affect the disease status of the individuals distributed among cases and controls. Such covariate information may be available in the form discrete data, categorical data or even in the form of continuous data (see, Lawless (2011), Agresti (2003)). The analyst tries to establish some sort of relationship among the covariate and the disease status of individuals so that the appropriate causal factor can be ascertained. These problems can be approached by using the recent Bayesian methods (see, Poole (2002), Martz and Waller (1982), Zelen and Parker (1986), Ashby et al. (1993), Ghosh and Chen (2002) etc.). Besides these, one can also undertake survival analysis in the medical data problems. In this case, the data are available in a continuous manner such as the time to occurrence of an event. This may be directly in the form of survival times, remission times or even in the form of time taken by a patient to get

* Corresponding author e-mail: rsrivastava.bhu@gmail.com

complete cure. In the recent developments for analyzing survival data one can refer Lin Zhang (2018) and Camille Maringe (2020). Apart from using single model there is also possibility of using competing risk survival models (see, Dominic Edelmann (2020)) Besides, we may have the situations where the experimenter faces with missing data problem. There may be situations when the availability of data is scarce often with ambiguous specification. Such type of situation is not easy to deal with unless some specific strategies developed by the analysts (Lawless (2011)), Martz and Waller (1982), Bossly (2020))

In present study, we propose statistical models and complete Bayesian inference for gallbladder cancer survival data of two countries. Gallbladder Carcinoma (GBC) is relatively uncommon cancer but very common in India. North-Eastern parts of India are more vulnerable to it. Among the various disease, cholelithiasis is the commonest, which requires surgical removal of the gallbladder of the patient, the treatment has high rate of success. The patients after surgical removal of their gallbladder maintain more or less normal lives. GBC, on the other hand involves life risk of the patients. It is the third most common gastrointestinal malignancy in this part of the country and three times more frequent among females than males. The factors responsible for this involves specific life style of the patients, reproductive factors, age factors, biological factors, typhoid carrier agents, heavy metals and gender of patients, amongst others (see, Kato et al. (1992), Chow et al. (1999), Dixit et al. (1998), Shukla et al. (1998), Pandey and Shukla (2002), Shukla et al. (2008)). Pandey and Shukla (2004) systematically describe the various gallbladder disease including GBC. Plenty of studies available on the etiological factors of gallbladder carcinoma as well as survival of cancer patients after radiotherapy and chemotherapy but very few involves the Bayesian concept of incorporating expert's opinion. Now-a-days bio-statisticians are focusing more on recent statistical methodologies and also trying to develop new and sophisticated methodologies to give more authenticated messages provided by medical studies.

The study is organized as follows. In next section, necessary modeling formulation for survival data has been discussed. We briefly describe Weibull and log normal distributions. Bayesian formulation for the propose models is discussed in section 3. Posterior complexity and Markov chain Monte Carlo (MCMC) simulations for the its computation has also been discussed in section 3. Latter in section 4, Bayesian tools for checking model compatibility and model comparison has been discussed in brief. In section 5, gallbladder cancer survival times has been considered for Indian as well as European context. Both sets of survival times has been analyzed considering Bayesian formulation discussed in section 3. This section has two subsections for checking model compatibility and model comparison for Bayesian survival models. Conclusion has been given in the last section.

2 Statistical Model for Survival Data

Considering the scenario of survival of the GBC patients, it can be facilitated and examined by using various life time, say for example, exponential, Weibull, generalized Weibull model(see, MC Jones (2019). MA Khan (2019)), Gamma, log-normal distributions (see Machin et al. (2006), Klein and Moeschberger (2013)). Among these models, Weibull and lognormal are perhaps most useful in lifetime data analysis. Probability density function of two parameter Weibull distribution is given by 1.

$$f_W(t) = \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1} \exp - \left(\frac{t}{\theta}\right)^{\beta} \tag{1}$$

where β is shape parameter of Weibull distribution while θ is scale parameter. Survival rate and mean life time (MLT) for Weibull distribution are given in 2 and 3

$$h_W(t) = \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1} \tag{2}$$

$$MLT_W = E(t) = \theta\Gamma\left(1+\frac{1}{\beta}\right) \tag{3}$$

where Γ denotes the gamma function. For $\beta < 1$ survival function $h_W(t)$ decreases with time, for $\beta > 1$ it increases with time and for $\beta = 1$ it reduces to constant $\frac{1}{\theta}$. There are other situations where the hazard rate could be increasing or decreasing with time. Say for example, the risk of death in a newborn infant is high in the initial days of life; thereafter it rapidly decreases and remains low and approximately constant, until much later in life, when it begins to increase again. Weibull distribution fits the data well in all three conditions.

Log-normal distribution is another widely used model in survival analysis. Applicability of log-normal distribution is perhaps due to simple structure and closeness to normal distribution. A non negative random variable t follows log-normal distribution if $\log(t)$ follows normal distribution. Probability density function of log-normal distribution is given in 4.

$$f_L(t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp - \left(\frac{\log t - \mu}{\sqrt{2}\sigma}\right)^2 \tag{4}$$

Survival function of log-normal distribution is not in close form and hence its survival rate is also not in closed form. MLT for log normal distribution is given by 5.

$$MLT_L = E(t) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \tag{5}$$

In contrast to that of Weibull distribution, this hazard function increases initially and then decreases with time. This form of hazard may arise in many clinical and epidemiological situations. For example, immediately after surgical removal of a cancer cell, the risk of tumor, recurrence is often small but then may increase for a period but thereafter begins to decline once more. It is observed that in many real life situations, the Weibull and log-normal distributions may fit a set of data equally well, especially over the central region of the distribution. When these models are fitted for a data set, the Weibull distribution has an earlier lower tail than the corresponding log-normal distribution, which indicates a low Weibull percentage of failures is below the corresponding log-normal ones. This is perhaps the reason that Weibull distribution is said to be more pessimistic than log-normal distribution (see Meeker and Escobar (2014)) though there are certain situations when the later model is recommended over the former. Aitchison and Brown (1966), Mann et al. (1974) and Upadhyay and Peshwani (2003) are some important references on both classical and Bayesian developments related to the log-normal distribution.

3 Bayesian Model Formulation and Gibbs Sampler

Bayesian analysis is marked by the subjective probability belief, or a prior description of the model parameter. It combines the prior distribution with the observed data via Bayes theorem to update our knowledge regarding the uncertainty and defines the posterior distribution. This posterior distribution is the basis of Bayesian inferences. Although posterior distribution is essentially an updated version of prior knowledge about model parameters in the light of observed data, prior distribution plays an important role in the Bayesian paradigm. It can be broadly classified as proper and improper priors, informative, non-informative priors (see,Martz and Waller (1982), Kass and Wasserman (1995), Berger (2013), Box and Tiao (2011)).

In this section, We are proposing Bayesian models for GBC survival data and algorithms for generating random sample from finally derived posterior for inferences and predictions. Since we do not have any prior information about parameters of both the models, we simply have used the non informative prior for both the case. These priors are very common and have been already verified by several authors (see, Yang and Berger (1996), Upadhyay et al. (2001)).

3.1 Bayesian Model Formulation for Weibull Model

Let $\underline{t} : t_1, ..., t_n$ are n survival times corresponding to Weibull model. Likelihood function for Weibull model can be written as

$$L(\underline{t}; \theta, \beta) = \prod_{i=1}^{n} f_W(t_i) = \frac{\beta^n}{\theta^{n\beta}} \prod_{i=1}^{n} (t_i)^{\beta-1} \exp - \sum_{i=1}^{n} \left(\frac{t_i}{\theta}\right)^{\beta}. \tag{6}$$

Let us consider the joint prior for parameters θ and β as

$$\pi(\theta, \beta) = \frac{1}{\theta\beta}. \tag{7}$$

This prior has been considered only because of unavailability of real expert. Combining likelihood (6) and prior (7) via Bayes theorem, joint posterior of θ and β up to proportionality can be written as (8)

$$\pi(\theta, \beta|\underline{t}) \propto \frac{\beta^{n-1}}{\theta^{n\beta+1}} \prod_{i=1}^{n} (t_i)^{\beta-1} \exp - \sum_{i=1}^{n} \left(\frac{t_i}{\theta}\right)^{\beta}. \tag{8}$$

Since posterior kernel in equation (8) is analytically intractable, posterior inferences can be drawn using sample based approaches in particular Gibbs sampler algorithm.

In order to implement Gibbs sampler algorithm, one need to specify full conditional distributions at least upto proportionality. Once these full conditionals are specified, posterior sample can be obtained by iteratively generating

sample from each of these full conditionals. Two full conditionals distribution corresponding to posterior (8) can be written as 9 and 10

$$\pi(\beta|\underline{t}) \propto \frac{\beta^{n-1}}{\theta^{n\beta+1}} \prod_{i=1}^{n} (t_i)^{\beta-1} \exp - \sum_{i=1}^{n} \left(\frac{t_i}{\theta}\right)^{\beta}. \tag{9}$$

$$\pi(\theta|\underline{t}) \propto \frac{1}{\theta^{n\beta+1}} \prod_{i=1}^{n} (t_i)^{\beta-1} \exp - \sum_{i=1}^{n} \left(\frac{t_i}{\theta}\right)^{\beta}. \tag{10}$$

Full conditional (9) can be shown log-concave and hence generation from it can be carried out using adaptive rejection sampling (see, Gilks and Wild (1992)). Considering transformation $\lambda = \theta^{\beta}$ in (10), λ can be shown to follow gamma distribution and can be generated using standard gamma subroutine.

Algorithm 1 Algorithm for generating from posterior surface equation (8)

Step 1: Take initial value $(\beta^{(1)}, \theta^{(1)})$
Step 2: Start for loop $k = 2, 3, \ldots, N$ where N is number of simulations
Step 3: Generate $\beta^{(k)}$ from $\pi(\beta|\underline{t}, \theta^{(k-1)})$ using adaptive rejection sampling
Step 4: Generate $\lambda^{(i)}$ from $\pi(\lambda|\underline{t}, \beta^{(k)}) = Gamma\left(Shape = n, Rate = \sum_{i=1}^{n} t_i^{\beta^{(k)}}\right)$
Step 5: Set $\theta^{(k)} = \{\lambda^{(k)}\}^{-1/\beta}$
Step 6: End for loop

Once posterior sample is obtained by above algorithm, one can easily get posterior summary simply by getting summary of this sample. One can also get various measures of model comparison using this sample.

3.2 Bayesian Model Formulation for Log-normal Model

Let $\underline{t} : t_1, \ldots, t_n$ are n survival times from log-normal model. Likelihood function corresponding to log-normal can be written as

$$L(\underline{t}; \mu, \sigma) = \prod_{i=1}^{n} f_L(t_i) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}} \prod_{i=1}^{n} t_i} \exp - \sum_{i=1}^{n} \left(\frac{\log t_i - \mu}{\sqrt{2}\sigma}\right)^2 \tag{11}$$

Joint prior for the parameter μ and σ can be given as

$$\pi(\mu, \sigma) = \frac{1}{\sigma}. \tag{12}$$

Combining likelihood (11) and prior (12) via Bayes theorem, joint posterior up to proportionality can be written as

$$\pi(\mu, \sigma|\underline{t}) = \prod_{i=1}^{n} f_L(t_i) = \frac{1}{(2\pi\sigma^2)^{\frac{n+1}{2}} \prod_{i=1}^{n} t_i} \exp - \sum_{i=1}^{n} \left(\frac{\log t_i - \mu}{\sqrt{2}\sigma}\right)^2 \tag{13}$$

Posterior surface in 13 is not available in closed form. So for generating sample from 13, sample based approaches i.e Gibbs sampler or Metropolis algorithm can be used. Gibbs sampler algorithm can be implemented in similar manner as discussed in previous section. Full conditionals for parameter μ and σ can be written as 14 and 15.

$$\pi(\mu|\underline{t}) \propto \exp - \sum_{i=1}^{n} \left(\frac{\log t_i - \mu}{\sqrt{2}\sigma}\right)^2 \tag{14}$$

and

$$\pi(\sigma|\underline{t}) \propto \frac{1}{(2\pi\sigma^2)^{\frac{n+1}{2}}} \exp - \sum_{i=1}^{n} \left(\frac{\log t_i - \mu}{\sqrt{2}\sigma}\right)^2 \tag{15}$$

respectively. Parameter μ can shown to follow a log-concave distribution. Hence, random sample for μ can be generated using adaptive rejection sampling scheme (see, Gilks and Wild (1992)). In full conditional 15 of parameter σ, transformation $\lambda_1 = 1/\sigma^2$ can be shown to follow Gamma distribution with scale parameter $\frac{2}{\sum_{i=1}^{n}(\ln x_i - \mu)^2}$ and shape parameter $n/2$. Hence σ can be generated using standard gamma subroutine. So, using these two generators, we can finally implement Gibbs sampler algorithm to obtain posterior samples.

Once posterior samples are obtained, one can easily get the desired posterior summary.

Algorithm 2 Algorithm for generating from posterior surface equation (13)

Step 1: Take initial value $(\mu^{(1)}, \sigma^{(1)})$
Step 2: Start for loop $k = 2,3,...,N$, where N is number of simulations
Step 3: Generate $\mu^{(k)}$ from $\pi(\mu|\underline{t}, \sigma^{(k-1)})$ using adaptive rejection sampling
Step 4: Generate $\lambda_1^{(k)}$ from $\pi(\lambda_1|\underline{t}, \mu^{(k)}) = Gamma\left(Shape = n/2, Scale = \frac{2}{\sum_{i=1}^{n}(\ln x_i - \mu^{(k)})^2}\right)$
Step 5: Set $\sigma^{(k)} = \sqrt{1/\lambda_1^{(k)}}$
Step 6: End for loop

Table 1: Gallbladder Carcinoma Survival time data-1(in days)

304.16	212.91	243.33	425.83	212.91	699.58	425.83	882.08
304.16	699.58	638.75	152.08	699.58	669.16	638.75	273.75
152.08	395.41	30.41	212.91				

4 Model compatibility and Comparison

Whenever we consider statistical model for any data, it is necessary to investigate whether model is compatible for the data. There are both formal and informal Bayesian approaches to investigate model compatibility. We are using latter approach in our study though reader can see Upadhyay et al. (2001). Informal approach is basically a graphical procedure where we plot any statistical function for both observed and predictive data. Same pattern of statistical function of data sets (observed and predictive) indicates the compatibility of model with data. One can consider hazard function, empirical cumulative distribution function (Cdf) plot etc as a choice of statistical function.
If two or more than two models found to be compatible for data in hand, next we have to investigate which of the model is more suitable to describe data among the group of compatible models. Information criterion is used for comparison of several compatible models. Among the information criterion, Bayes information criterion (BIC) and the Deviance information criterion (DIC) have been considered extensively in the literature. The BIC for model i can be defined as

$$BIC_i = -2x\log(\hat{L}_i) + m_i x\log(n), \qquad i = 1,2,... \tag{16}$$

where $\hat{L}_i$ is the value likelihood at posterior mean. m_i is the number of parameters involved in the model i and n is number of observations in data set. The model corresponding to lowest value of BIC is recommended.
DIC is another good criterion for comparing the two or more models for the given dataset. This criterion was suggested by Spiegelhalter et al. (2002). An important advantage of DIC is that it can be calculated using posterior sample. DIC for model i can be defined as

$$DIC_i = \overline{D_i(\Theta_i)} + p_{Di}, \qquad i = 1,2,... \tag{17}$$

where $D_i(\Theta_i)$ is the deviance defined as $D_i(\Theta_i) = -2\log(L_i(\underline{Y}|\Theta_i))$, where $\overline{D_i(\Theta_i)}$ is the posterior mean of deviance, Θ_i denotes the parameter vector associated with the model i. The second term p_{Di} in equation is the number of effective parameters for the model i, defined as $p_{Di} = \overline{D_i(\Theta_i)} - D(\hat{\Theta}_i)$, where $\hat{\Theta}_i$ is an estimate of Θ_i, usually taken as posterior mean. The term p_{Di} can also be interpreted as a measure of model complexity.
In present article, we have used these two criterion to compare the Weibull and lognormal model for the given data set.

5 Bayesian survival analysis of gallbladder carcinoma patients

Two sets of survival times of gallbladder cancer patients has been considered. We are indicating these data sets as data1 and data2 (see Table- 1 and Table-2). Data1 is with reference to Indian context reported in Doval et al. (2014) whereas data2 reported in Bruha et al. (2007), is in European Context. We could not arrange the complete data sheet reported in Doval et al. (2014). Although both data sets consist of survival times of gallbladder cancer patients, we can not analyze it by taking together because of there space differences. Environmental conditions, medical facilities and many other factors of both the regions are different and these factors may affect the survival times. Both data sets has been analyzed separately in Bayesian context by considering Weibull as well as log-normal models in each case. We are also providing Bayesian model comparison in Indian and European context.

Table 2: Gallbladder Carcinoma Survival time data-2(in days)

86	92	127	137	176	180	189	192
210	213	225	234	397	609		

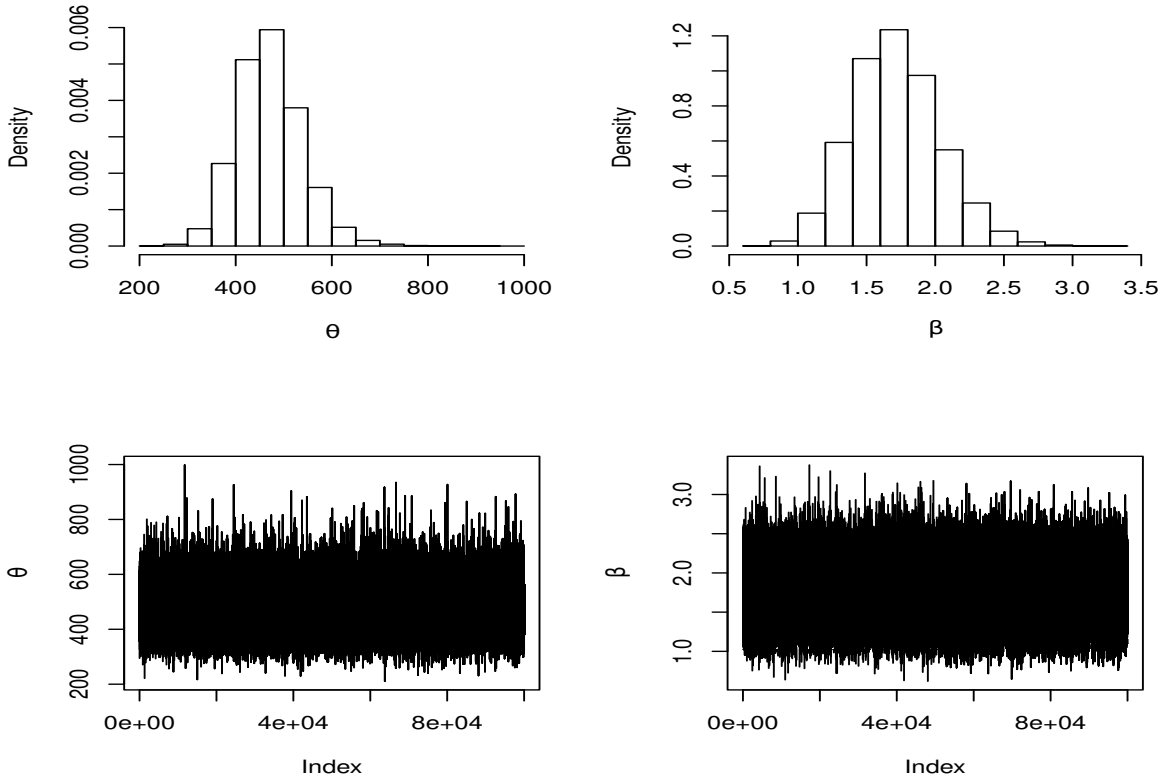

Fig. 1: Posterior density of parameters of Weibull model for data 1

Model formulation of survival data is done by considering two life time models discussed in Section 2. Further we consider non-informative prior in order achieve Bayesian formulation of the problem(see, Section 3). We have analyzed both models (1 and 4) for both data sets using Bayesian paradigm. In order to get posterior summary, we generated random sample from posterior surface (8) and (13). For the same we implemented Gibbs sampler algorithm given in subsection 3.1. we have simulated 50K observations from posterior (8) and (13) using R software-programming, corresponding both data sets. In each of four cases, convergence has been obtained after 10K iterations. Hence, we discarded first 10K observation due to burn in phase and selected 1K independent observation by taking gap of 40 remove serial correlation. Marginal posterior density estimates of parameters of Weibull model i.e. θ, β for data1 and data2 are shown in Figure (1) and Figure (2), respectively, in form of histograms. Marginal posterior density estimates of μ, σ of lognormal model are presented in Figure (3) and Figure (4), respectively.

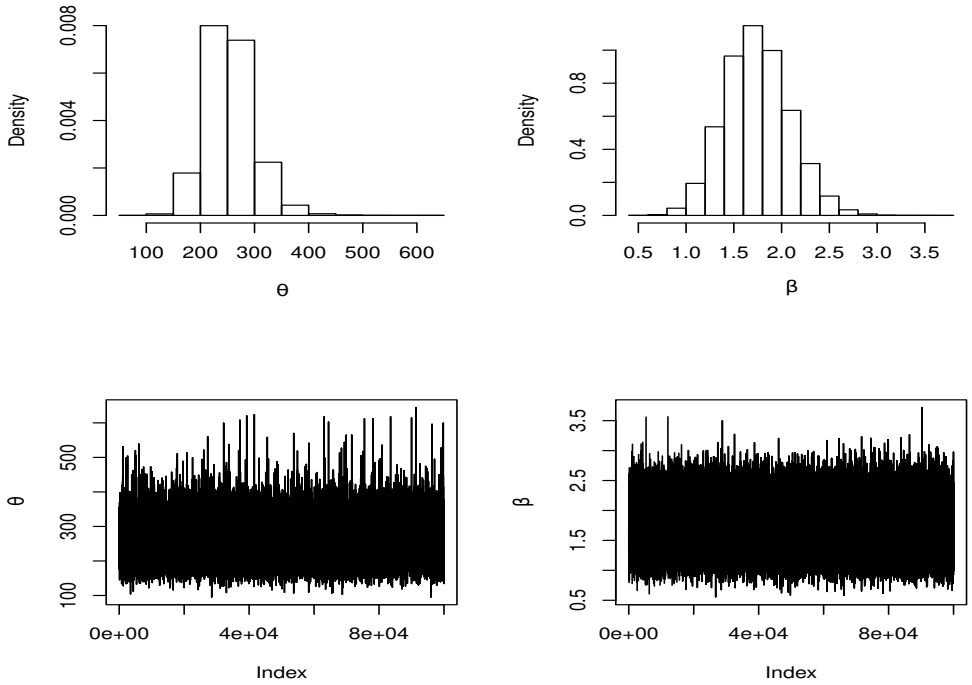

Fig. 2: Posterior density of parameters of Weibull model for data 2

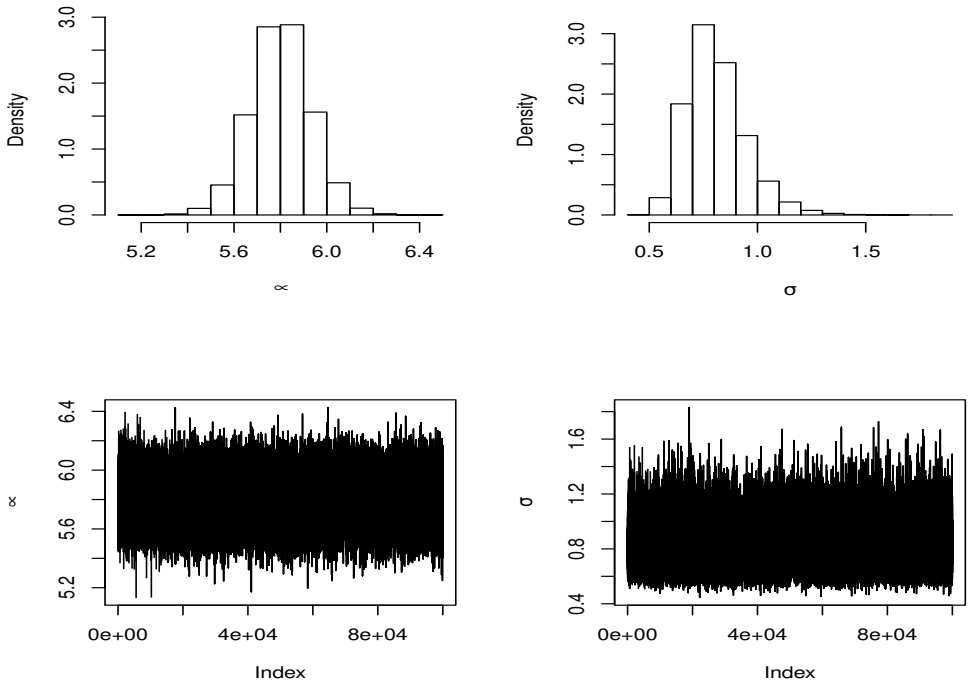

Fig. 3: Posterior density of parameters of lognormal model for data 1

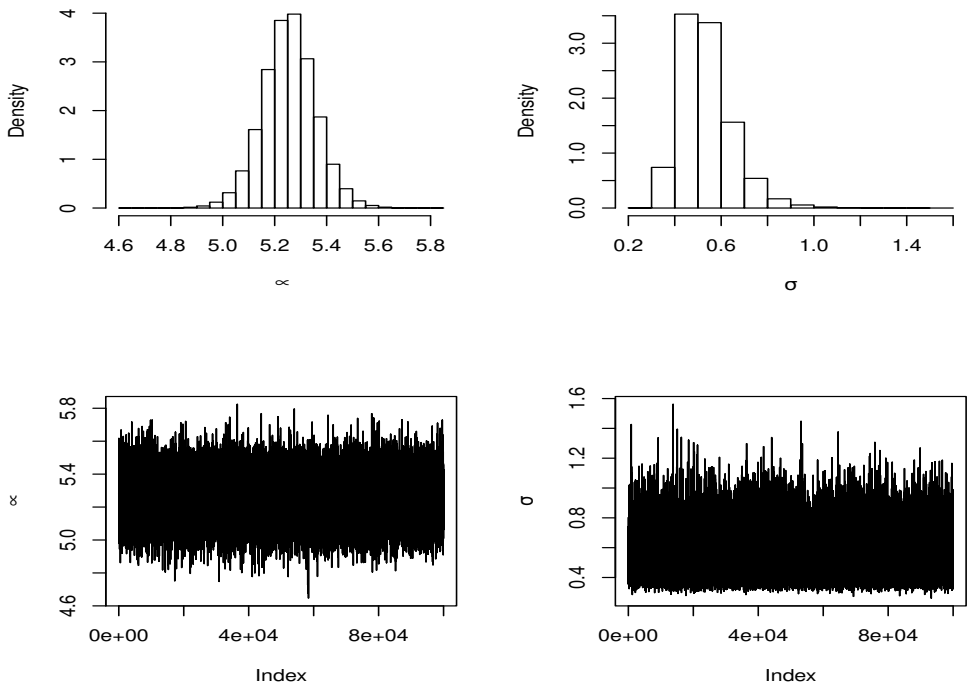

Fig. 4: Posterior density of parameters of lognormal model for data 2

In each four cases, based on these 1K observation, we have obtained the posterior summary which is given in Table 3 and Table4.

Table 3: Posterior characteristic of parameters of Weibull model for both data sets

Considered data	Variates	Estimated posterior Mean	Median	Mode	Estimated 0.95-*HPD*	
data1	θ	470.624	466.857	464.270	342.076	607.431
	β	1.710	1.704	1.679	1.126	2.365
data2	θ	254.388	251.195	242.928	172.274	343.594
	β	1.745	1.732	1.691	1.094	2.421

Table 4: Posterior characteristic of parameters of lognormal model for both data sets

Considered data	Variates	Estimated posterior Mean	Median	Mode	Estimated 0.95-*HPD*	
data1	μ	5.804	5.804	5.799	5.531	6.046
	σ	0.808	0.791	0.766	0.581	1.092
data2	μ	5.256	5.257	5.267	5.044	5.449
	σ	0.533	0.516	0.481	0.346	0.752

Posterior summary for Weibull model corresponding to both data sets are given in Table 3. Posterior mean, median, mode and 95%*HPD* interval has been estimated to summarize . Similar results corresponding log normal model for both data sets are given in Table 4.

Besides posterior summary, posterior average of mean lifetime for both Weibull model and log normal(see, equation (3) and (5)) has also been obtained. We have also obtained the 95%*HPD* interval of mean lifetime for both models and for both data sets. These results are given in Table 5.

Table 5: Posterior average for mean lifetime

Model	Data	Mean	0.95 HPD interval
Weibull	data1	422.14	(307.99 542.90)
	data2	228.58	(155.44 305.64)
Log normal	data1	470.54	(329.39 643.604)
	data2	223.99	(176.05 279.70)

Since we have very small data sets and non informative priors for model (1) and (4), we can not compare the mean lifetime of two datasets. Observed mean lifetime for data1 and data2 are 413.66 and 219.07 respectively.

5.1 Model Compatibility based on Real Data Set

As mentioned in Section 4, model compatibility using observed and predictive empirical Cdf plots corresponding to each data set and each model has been performed in this study. These plots are shown in Figure (5-8). In every case, first we have plotted the empirical Cdf of actual observations, then we have plotted the empirical Cdf plot based posterior predictive observations corresponding to the considered model. Since we have two data sets and two models, we have four plots in total.

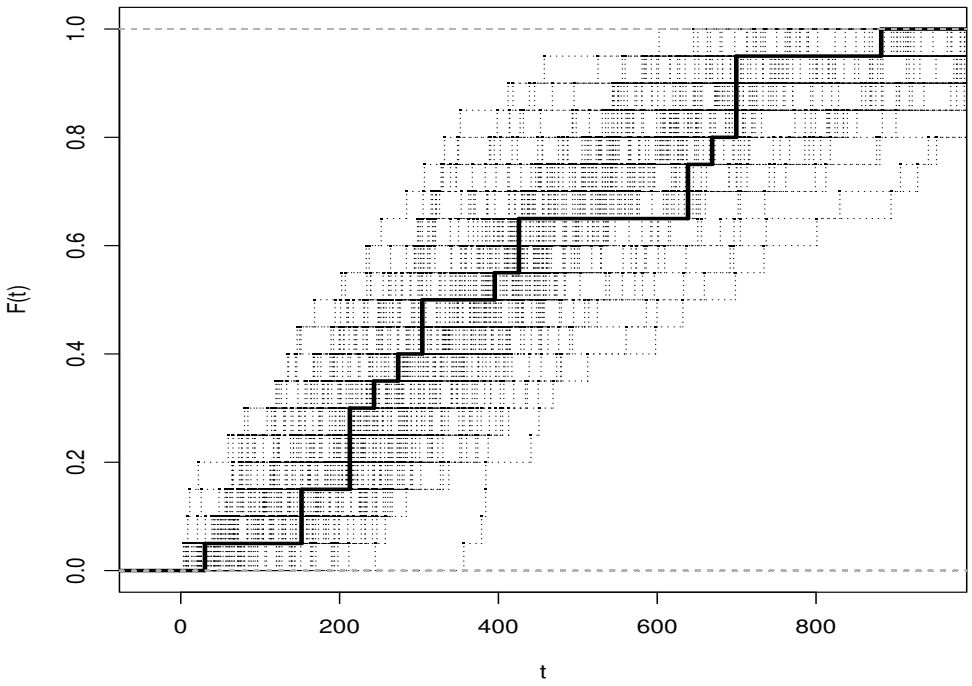

Fig. 5: Observed and predictive Cdf Plot of Weibull model for data1

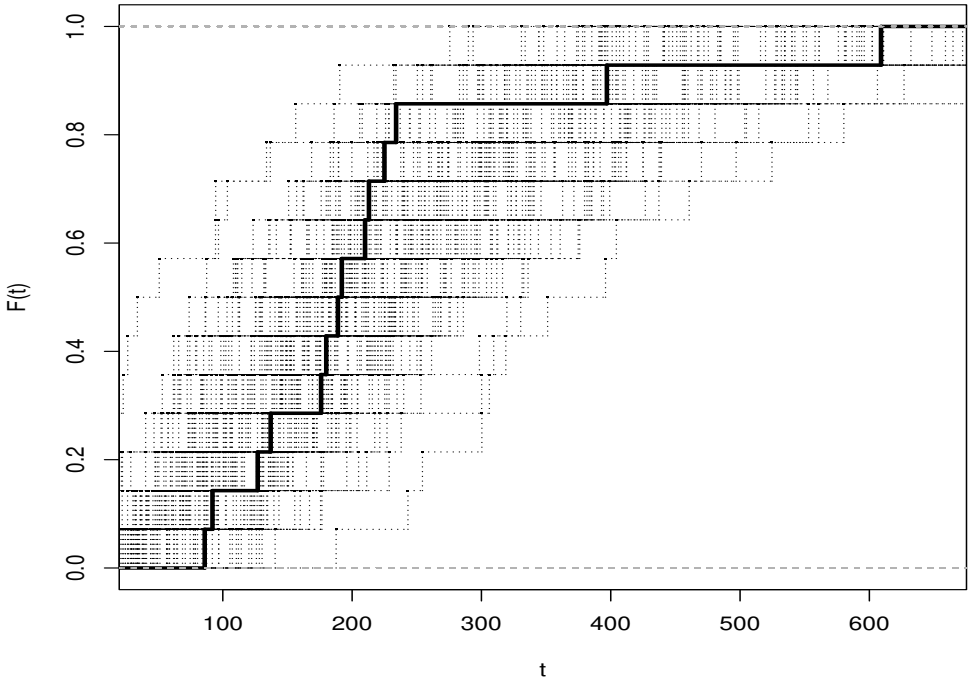

Fig. 6: Observed and predictive Cdf Plot of Weibull model for data2

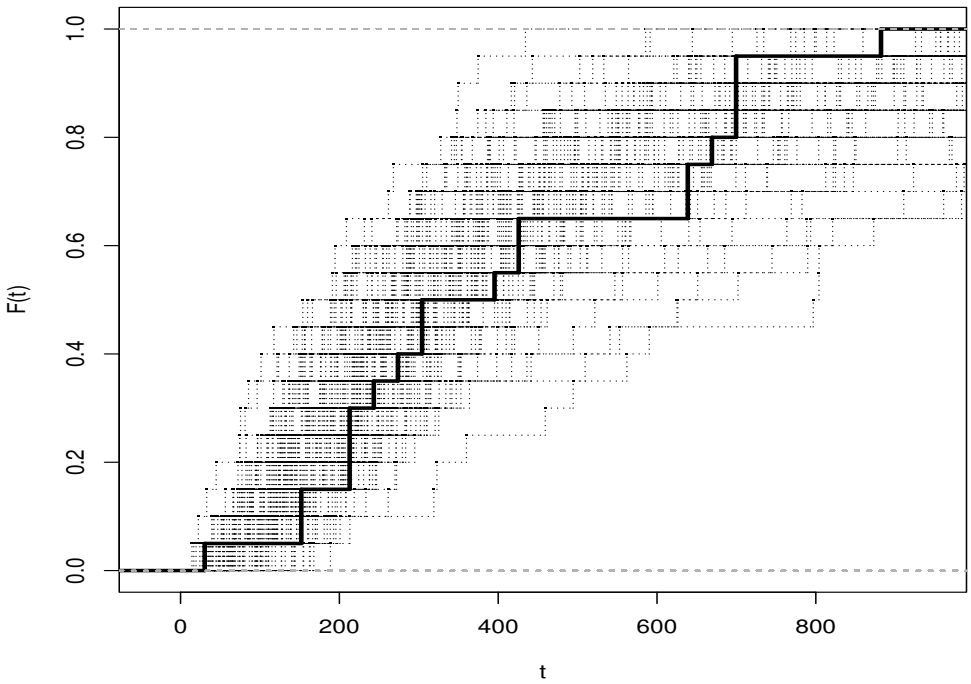

Fig. 7: Observed and predictive Cdf Plot of Lognormal model for data1

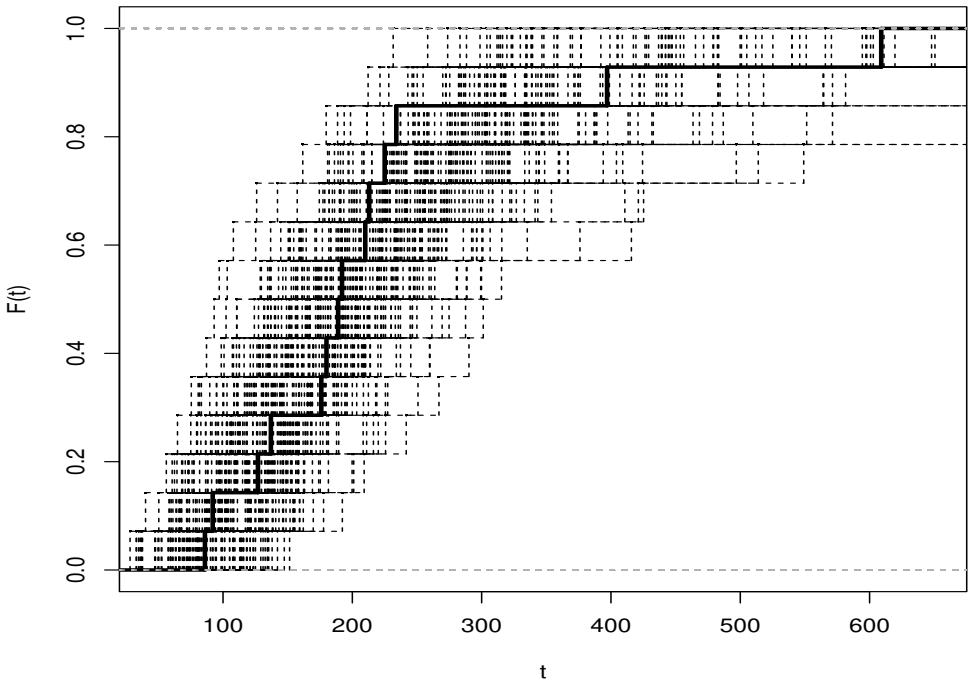

Fig. 8: Observed and predictive Cdf Plot of Lognormal model for data2

Looking at each plots, we can say that both models are fitted well to both data sets. So, finally, at this stage, we are not in position to give any conclusion about the best model based on these plots. Here, we can only recommend model comparison study for both model based on two datasets. Next subsection provides the model comparison based on the real data sets.

5.2 Model comparison for real data set

We have compared lognormal and Weibull model for both data sets using Bayesian model comparison methods mentioned in Section 4. Result of model comparison is shown in Table 6. BIC is calculated based on maximized value of log likelihood while DIC is calculated using posterior sample of size 1000 as mentioned earlier.

Table 6: Model comparison table for both data set.

Data	data1		data2	
Model	Weibull	Lognormal	Weibull	Lognormal
BIC	313.0245	318.27	199.52	195.18
DIC	277.1208	281.21	175.70	170.18

Based on result provided in Table 6, we can conclude that Weibull model gives better fit than log normal for survival time of gallbladder carcinoma patients in Indian context. In European context, the log normal model gives better fit than Weibull model. Perhaps, this indicate that European patient suffering with carcinoma have a good survival rate after the treatment.

6 Conclusion

Model formulation of medical data plays very important role in medicine and health care sector for making prediction, providing treatments, identifying effect of various therapies etc. Also in medical field one is compel to take observations, coming by their own from human being as one cannot perform experiments on humans. This limitation can be tackle

by taking information from medical expert of the field and use this information as prior information of the scenario. Further Bayesian paradigm enables to incorporate prior knowledge about the model parameter with the observed data through Bayes theorem and provides posterior distribution for further analysis. Posterior distribution has the essence of both prior information as well as observed data. This study is a successful attempt of complete Bayesian analysis of two survival models and proposes appropriate statistical models for survival times of gallbladder cancer patients in Indian and European context. Compatibility of both models for both contexts has been checked and selection of the most suitable model is done by using measures of information criterion. In this study, Weibull model becomes an appropriate modelling assumption in Indian context whereas lognormal fits well for the data in European context. Better fitting of Weibull model in Indian context may indicate an increasing trend of hazard rate in Indian patients whereas fitting of lognormal in other context indicates initial increase and then decreasing trend of hazard rate in European context. This is perhaps because of good medical facility in Europe. Also environmental conditions, medical facilities and many other factors of both regions are different and these factors affect the survival times. Proposed study is also applicable to understand the time to the development of diabetic retinopathy from time to diagnosis of diabeties.

Acknowledgment

The research work of Dr Rakesh Ranjan is supported by IoE Scheme of Banaras Hindu University.The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Agresti, A. (2003). Logit models for multinomial responses. *Categorical Data Analysis, Second Edition*, pages 267–313.

Aitchison, J. and Brown, J. (1966). *The Lognormal Distribution: With Special Reference to Its Uses in Economics*. Cambridge. University. Dept. of Applied Economics. Monographs, 5. University Press.

Ashby, D., Hutton, J. L., and McGee, M. A. (1993). Simple bayesian analyses for case-control studies in cancer epidemiology. *The Statistician*, pages 385–397.

Bassetti, W., Woodward, M., and Woodward, M. (2004). *Epidemiology:Study Design and Data Analysis, 2nd ed.* Chapman & Hall/CRC.

Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.

Bossly, A. A. (2020). Bayesian statistics applications on reliability prediction and analysis. *Journal of Statistics Applications & Probability*, 9(1):19–34.

Box, G. E. and Tiao, G. C. (2011). *Bayesian inference in statistical analysis*, volume 40. John Wiley & Sons.

Bruha, R., Petrtyl, J., Kubecova, M., Urbanek, P., and Marecek, Z. (2007). *Metallic Selfexpandable Stents Combined with Intraluminal Brachytherapy in the Treatment of nonresectable biliary malignancies*, chapter 6, pages 57–73. Horizons in cancer research. Nova Science Publishers, New York.

Camille Maringe, Aurélien Belot, B. R. (2020). Prediction of cancer survival for cohorts of patients most recently diagnosed using multi-model inference. *Statistical Methods in Medical Research*, 29(12):3605–36227.

Chow, W., Johansen, C., Gridley, G., Mellemkjaer, L., Olsen, J., and Fraumeni, J. (1999). Gallstones, cholecystectomy and risk of cancers of the liver, biliary tract and pancreas. *British Journal of Cancer*, 79:640–644.

Dixit, V. et al.. (1998). Xanthogranulomatous cholecystitis. Dig Dis Science, 43:940–942. Xanthogranulomatous cholecystitis. *Dig Dis Science*, 43:940–942.

Dominic Edelmann, Maral Saadati, H. P. J. G. (2020). A global test for competing risks survival analysis. *Statistical Methods in Medical Research*, 29(12):3666–3683.

Doval, D. C., Azam, S., Sinha, R., Batra, U., and Mehta, A. (2014). Expression of epidermal growth factor receptor, p53, bcl2, vascular endothelial growth factor, cyclooxygenase-2, cyclin d1, human epidermal receptor-2 and ki-67: Association with clinicopathological profiles and outcomes in gallbladder carcinoma. *Journal of carcinogenesis*, 13:1–10.

Ghosh, M. and Chen, M.-H. (2002). Bayesian inference for matched case-control studies. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 107–127.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for gibbs sampling. *Applied Statistics*, pages 337–348.

Kass, R. E. and Wasserman, L. (1995). A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. *Journal of the american statistical association*, 90(431):928–934.
Kato, I., Nomura, A., Stemmermann, G., and Chyou, P. (1992). Prospective study of the association of alcohol with cancer of the upper aerodigestive tract and other sites. *Cancer Causes Control*, 3:145–151.
Klein, J. and Moeschberger, M. (2013). *Survival Analysis: Techniques for Censored and Truncated Data*. Statistics for Biology and Health. Springer New York.
Lawless, J. F. (2011). *Statistical models and methods for lifetime data*, volume 362. John Wiley & Sons.
Lin Zhang, I. K. (2018). Semiparametric bayesian kernel survival model for evaluating pathway effects. *Statistical Methods in Medical Research*, 28(10-11):3301–3317.
Khan, M. A. and Srivastava, R. (2019). New modified weibull distribution: moment properties and bayes analysis. *Journal of Statistics Applications & Probability*, 8(3):1–8.
Machin, D., Cheung, Y., and Parmar, M. (2006). *Survival Analysis: A Practical Approach*. Wiley.
Mann, N. R., Singpurwalla, N. D., and Schafer, R. E. (1974). *Methods for statistical analysis of reliability and life data*. Wiley, New York.
Martz, H. F. and Waller, R. A. (1982). *Bayesian Reliability Analysis.* Wiley, New York.
MC Jones, Angela Noufaily, K. B. (2019). A bivariate power generalized weibull distribution: A flexible parametric model for survival analysis. *Statistical Methods in Medical Research*, 29(8):2295–2306.
Meeker, W. Q. and Escobar, L. A. (2014). *Statistical methods for reliability data*. John Wiley & Sons.
O'Hagan, A. (1998). Eliciting expert beliefs in substantial practical applications. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1):21–35.
Pandey, M. and Shukla, V. (2002). Diet and gallbladder cancer: a case–control study. *European Journal of Cancer Prevention*, 11(4):365–368.
Pandey, M. and Shukla, V. (2004). *Gallbladder Cancer.* Jaypee, New Delhi.
Poole, D. (2002). Bayesian estimation of survival from mark-recapture data. *Journal of Agricultural, Biological and Environmental Statistics*, 7:264.
Shukla, V., Chauhan, V., Mishra, R., and Basu, S. (2008). Life style, reproductive factors and risk of gallbladder cancer. *Singapore Med J*, 49(11):912.
Shukla, V., Prakash, A., Tripathi, B., Reddy, D., and Singh, S. (1998). Biliary heavy metal concentrations in carcinoma of the gallbladder: Case control study. *British Medical Journal*, 317:1288–1289.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.
Srivastava, R., Upadhyay, S. K., and Shukla, V. (2018). Subjective elicitation of hyperparameters of a conjugate dirichlet prior and the corresponding bayes analysis. *Communications in Statistics - Theory and Methods*, 48(19):4874–4887.
Upadhyay, S. and Peshwani, M. (2003). Choice between weibull and lognormal models: A simulation based bayesian study. *Communications in Statistics-Theory and Methods*, 32(2):381–405.
Upadhyay, S., Vasishta, N., and Smith, A. (2001). Bayes inference in life testing and reliability via markov chain monte carlo simulation. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 63(1):15–40.
Yang, R. and Berger, J. O. (1996). *A catalog of noninformative priors*. Institute of Statistics and Decision Sciences, Duke University.
Zelen, M. and Parker, R. (1986). Case-control studies and bayesian inference. *Statistics in Medicine*, 5(3):261–269.